Quantification of Trace Chemicals using Vehicle Cabin Atmosphere Monitor

Seungwon Lee, Lukas Mandrake, Benjamin Bornstein, Brian Bue
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive Pasadena CA 91109
Seungwon.Lee@jpl.nasa.gov

Abstract— A system to monitor the concentrations of trace chemicals in cabin atmosphere is one of the most critical components in long-duration human flight missions. The Vehicle Cabin Atmosphere Monitor (VCAM) is a miniature gas chromatograph mass spectrometer system to be used to detect and quantify trace chemicals in the International Space Station. We developed an autonomous computational process to quantify trace chemicals for use in VCAM. The process involves the design of a measured signal quantification scheme, the construction of concentration curves (i.e. the relationship between concentration and ion count measured by VCAM), the decision rule of applying high- or low-gain concentration curves, and the detection of saturation, low-signals, and outliers. When the developed quantification process is applied, the average errors of concentration for most of trace chemicals are found to be between 14% and 66%. [1] [2]

TABLE OF CONTENTS

1. INTRODUCTION 1
2. APPROACH 1
3. RESULTS 4
4. CONCLUSIONS 5
REFERENCES 6
BIOGRAPHY 12

1. INTRODUCTION

The Vehicle Cabin Atmosphere Monitor (VCAM) is a gas chromatograph mass spectrometer instrument and is designed to autonomously identify trace organic species in the International Space Station (ISS) internal air and to quantify their concentration [1,2]. VCAM uses a gas chromatograph to separate chemicals in terms of time of arrival to the mass spectrometer. The chemicals travel through a specially prepared glass tube (column) in the gas state. The interactions of these gaseous analytes with the walls of the column causes different compounds to emerge at different times which are called retention times. After the chemicals are time separated, VCAM uses a quadrupole ion trap mass spectrometer to make unique mass fractionation patterns of each chemical analyte by applying a quadrupole RF electric field to hyperbolic electrodes and ramping the amplitude of the electric field at a constant rate [3].

Using the two instruments together, VCAM is designed to identify and quantify trace organic species. One of the operation requirements for VCAM is the autonomous quantification of the concentrations for approximately forty trace species within 40% error for a range of concentrations guided by the Spacecraft Maximum Allowable Concentrations (SMACs). In this paper, we present the autonomous process that we developed to quantify the concentrations of the trace chemicals.

2. APPROACH

VCAM divides the trace gas species to be identified and quantified into three priorities: Priority 1, Priority 2, and Priority 3. Tables 1, 2, and 3 list these groups and their desired detection concentration ranges. Priority 1 has the requirement of one hundred percent successful detection and identification. Priority 2 and Priority 3 require 80% and 70% detection and identification, respectively. For all three priorities, the error budget for quantification is 40%.

Designing Signal Quantification Scheme

In order to quantify the concentrations of the trace species, we first must design a scheme to quantify the measured signal of an elution peak. We examined three different ways to quantify the signal and selected the best performing scheme.

The first approach is to use the total ion count (TIC) under an elution peak subtracting an estimated background signal as shown in Figure 1. The total ion count is the sum of all the ion counts that are measured during the elution peak regardless of the mass channel numbers of the mass spectrometer. As a result, the total ion count will necessarily include contributions from noise or other nearby events.

The second approach is to use the sum of the ion counts of the mass channels only from perceived "data" mass peaks based on which mass peaks contributed most greatly to the elution peak's total height. We call this quantity data ion count (DIC). This approach directly couples the peak detection performed by the NIST Automated Mass Spectral Deconvolution and Identification System (AMDIS) algorithm [4,5].

Figure 2 shows a typical mass spectrum of an elution peak. The total ion count under an elution peak is distributed

1

[1] 978-1-4244-2622-5/09/$25.00 ©2009 IEEE.
[2] IEEEAC paper #1273, Version 6, Updated January 5, 2009

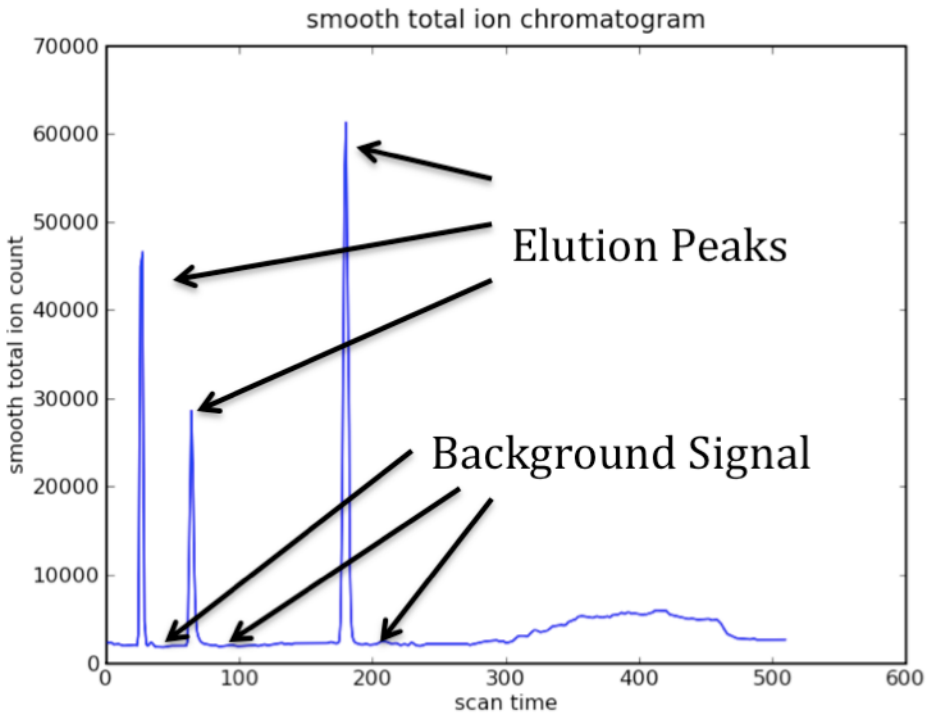


Figure 1 – Total Ion Chromatogram measured by VCAM Test-Bed (TB) Unit.

among different mass values depending on the ionized fragmentation of a given chemical. DIC is the sum of some of the red bars that AMDIS classifies as a major contributing mass peak for the given elution peak. There are several mass peaks that are persistent across elution and background frames for an entire run. These peaks appear due to the imperfect vacuum of the system. AMDIS scheme can distinguish the persistent mass peaks from signal mass peaks that truly belong to an elution peak. This method also attempts to distinguish between nearby elution events than may be partially merged (coeluted) in time.

The third approach is to use the sum of the ion counts of the mass channels specified by the library record identified as most likely by AMDIS. We call this quantity library ion count (LIC). The library mass peaks are the mass peaks that should appear for a given compound due to its ionized fragmentation pattern while it undergoes an ionization process. In principle, data mass peaks should be identical to library mass peaks for a given successfully identified event. But in practice data mass peaks often contain certain extra peaks due to system contamination, signal saturation, noise, or coelutions, while some expected library peaks may be missing due to sensitivity issues. Further, the NIST spectral library was defined relative to specific hardware which will never precisely match any other given configuration for all compound fragmentation patterns.

Figure 2 shows a comparison between the measured spectrum and the library spectrum for several chemicals. The fluorobenzene mass spectrum shows the typical saturation distortion pattern resulting from high chemical concentration measured with high-gain mode. The broader peaks in the measured spectrum will lead to the LIC being much smaller than the DIC due to the detected mass peaks far exceeding the library mass peaks in number. In contrast, the perfluoropropane mass spectrum shows a typical weak signal distortion pattern when low chemical concentration is measured with low-gain mode. Several mass peaks at high masses are missing or very weak compared to the expected library spectrum. The m-xylene spectrum shows an ideal situation where the measured spectrum looks very similar to the library spectrum across all mass peaks. The extra peaks around mass 30 are persistent mass peaks; the AMDIS-based mass peak finding algorithm (DIC) will ignore such contributions. For this ideal case, DIC is very similar to LIC.

The three different ways of defining a signal will give similar results, when the system is clean, the signal is neither saturated nor weak. However for a special case where we have coelution peaks that have large overlaps of the ion counts of multiple compounds, we should decompose the ion counts into several groups of ion counts for each different compound. LIC provide a means to decompose the ion counts for the coelution peaks if the library mass spectrum of the coeluting chemicals are largely orthogonal. Unfortunately, this is not the case for the majority of chemicals within our library which have similar retention times.

Constructing Concentration Curves

After the measured signal intensity is quantified in terms of integrated number of ion counted, we construct concentration curves. These curves capture the relationship between quantified signal counts and the physical concentration of the eluting chemical. After trying several different functional forms, the concentration curves are empirically defined as the log-log linear equation:

$$\log_{10}(\text{concentration}) = a\log_{10}(\text{signal}) + b \quad (1)$$

We obtain the optimal value for the slope a and the intercept b using a chi-square fitting for each chemical separately. Training data were weighted according to measurement errors and intrinsic measurement fluctuations. Since we have duplicate measurements for the same concentration, we can define the measurement fluctuation with the standard deviation of the quantified signal. We initially tried a non-weighted least square fitting method but found that the method failed when data contain many outliers and fluctuations. Chi-square fitting was found to be less subject to such anomalies and noise than a non-weighted least square fitting because the chi-square fitting take into account the measurement fluctuation and give a lower weight on the measurements with higher fluctuations. When the outliers and inhomogeneous fluctuations are manually removed, both the chi-square fitting and non-weighted least square fitting led to a similar concentration curve.

Gain Switch Decision Rule

VCAM performs in a dual-gain mode (high and low) in order to increase its dynamic range of concentration sensitivity. The gain determines the signal strength by varying the pulse duration of the ionization process. A higher gain means a longer pulse-duration, more ionization, and a higher counts for a given concentration. The reason that we use the dual-gain mode is that VCAM has the wide desired detection concentration range of the compounds so that one fixed gain cannot satisfy the concentration range. For example, if the gain is too low, we cannot detect a low concentration chemical. Conversely, if the gain is too high, we will have a saturated signal for a high concentration chemical so that the signal does not grow any further even though the concentration increases. As a result, we cannot estimate the concentration correctly. Having the dual-gain model addresses the shortcoming of the one-gain problems. The high-gain mode provides a sufficient signal for a low concentration chemical, while the low-gain mode provides

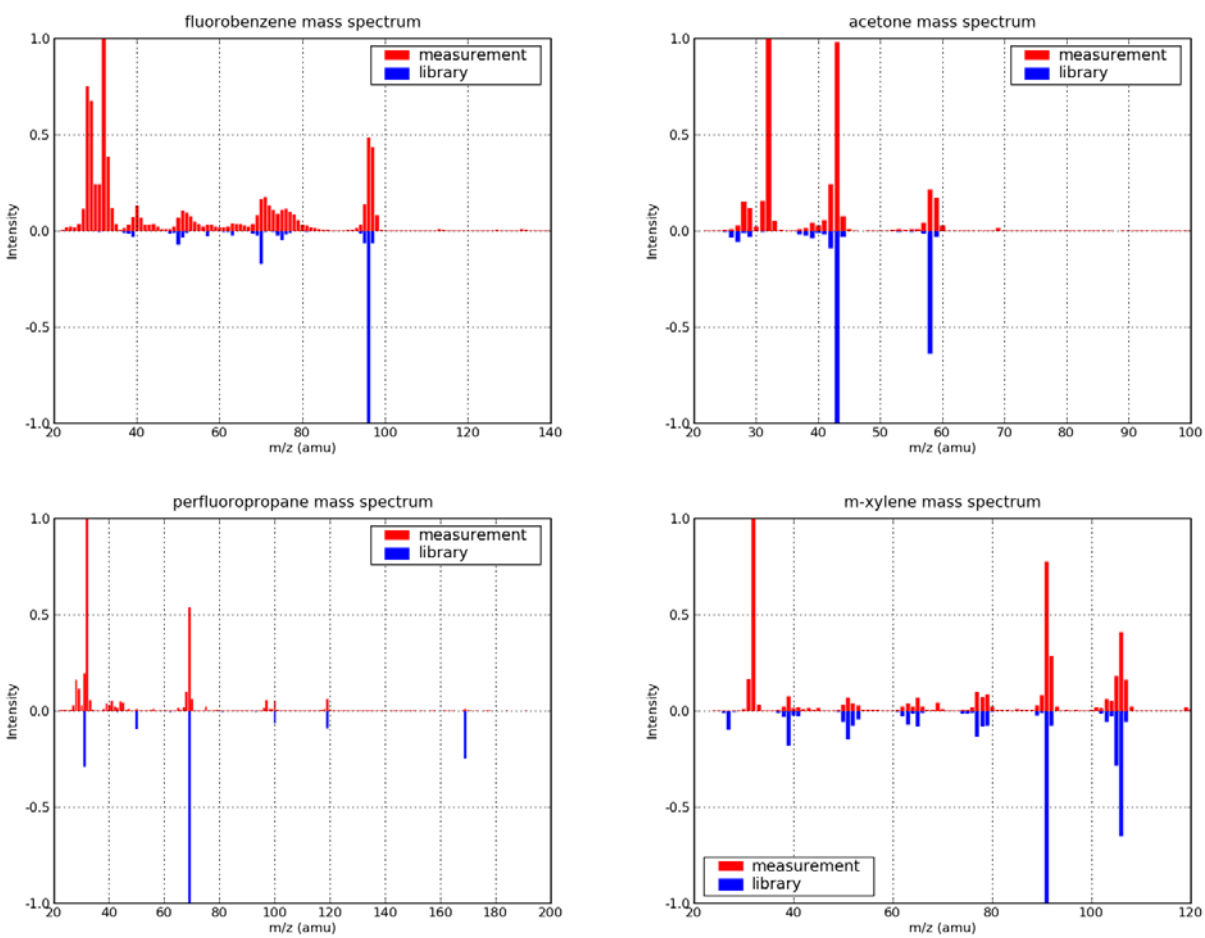


Figure 2 - Mass spectra of several chemical species. Each compound has a different mass spectrum with different mass peaks and different relative intensities of the peaks. DIC is calculated by adding the ion counts of mass peaks shown in the measured spectrum while LIC is calculated by adding the ion counts of mass peaks that coincide with library mass peaks in the library spectrum.

an unsaturated signal for a high concentration. Both high and low gain data are simultaneously taken for every run by alternating a long and a short pulse duration.

Although the dual gain does permit unsaturated detection of all specified ranges, we must decide when to use the high or low gain data without prior knowledge of sample concentration. Since both gain modes have different signal levels, we first must construct a concentration curve for each mode separately. After the concentration curves for both high and low gains are established, we have a choice of which gain-mode concentration curves to use for a detected signal. Since we have the dual-gain mode, we can select one mode that is more reliable than the other and use the mode to estimate the concentration. When a saturated signal is detected, this means that the high-gain signal is saturated so we will use a low-gain mode signal, apply the low-gain concentration curve, and report the concentration. Conversely, when a low signal is detected, this means that the low-gain signal is too weak to use so we will use a high-gain signal to determine the concentration.

This gain switch decision rule relies on the algorithm to determine whether the signal is saturated or weak. We developed a simple empirical algorithm for the determination of the saturated or weak signal. We take the maximum number of ion counts for the highest peak line above mass 33 in order to avoid the persistent N_2 and O_2 peaks at 28 and 32, respectively. If the maximum ion count is more than 8000, it is considered saturated. If the maximum ion count is less than 300, it is considered weak. Applying this rule leads to the choice of a more reliable gain in case of the existence of a saturated signal or a weak signal.

3. RESULTS

Selecting Best Signal Quantification Scheme

In order to establish concentration curves, we measured over 1100 elution events including over 30 chemicals, 5 different concentration values per chemical, and about 3 duplicates for each chemical and concentration. Figure 3 shows the concentration curves of the chemical species that we tested on the VCAM lab standard unit using three different ways to quantify a signal (TIC, DIC, and LIC). The symbols represent the quantified values of the measured signal and the lines represent the fitted concentration curves using the measurements.

As shown in Figure 3, we observed that DIC generates many more outliers that the other two ion-count methods (TIC and LIC). The large fluctuation of DIC for the same concentration is due to the sensitivity of our mass-peak finding algorithm to the signal fluctuation. The peak finding algorithm determines whether some signal is a contributor to a particular elution peak based on the signal sharpness relative to the noise/background level. This can vary run to run resulting in differing number of mass peaks considered contributing. Although this fluctuation is relatively harmless to identification, the DIC concentration scheme relies on these peaks for its count integration and is distored by such run-to-run noise. For this reason, we decided to not use DIC as the ion-count determining method for concentration curves.

Between TIC and LIC, the fitting error of the concentration curve is usually smaller with TIC for most of chemicals. One disadvantage of using LIC is its sensitivity to mass calibration errors and the dissimilarity of the library spectrum to the real spectrum. Our system experienced mass calibration errors of up to 1 AMU in certain spectral ranges, resulting the misalignment of entire significant peaks. Similarity of the library spectrum to the observed spectrum is also required for LIC to make sense, a requirement which held true approximately 75% of the time in our data. On the other hand, LIC has a strong advantage over TIC in terms of its ability to decompose the ion count from multiple chemicals in the case of coelution. However, the decomposition is not perfect unless the chemicals in the coelution peak have distinguishable library spectra. Most of our coeluting chemicals have similar library spectra, which complicate the ion count decomposition of the involved chemicals. Therefore, we selected TIC as the signal quantification method for our concentration curves although we are aware of its inability of decomposing the ion count under a coelution peak. With more accurate mass calibration and the use of the customized spectral library in future systems, LIC is predicted to be the most reliable method.

Obtaining Concentration Curves

Figure 3 shows that the weighted chi-square fitting is more reliable in extracting the overall trend without being obscured by outliers. When there are strong outliers (mostly in DIC case), the nonweighted least square fitting can be wildly affected by the strong outliers. Therefore, we decided to use the weighted chi-square fitting to establish the concentration curves.

Tables 2 and 3 show the resulting concentration curves of the required chemical species for the high and low gain using the chi-square fitting and TIC as the signal quantification method. The slope of the log-log curve (a) is typically between 1 and 3. The intercept of the log-log curve (b) varies considerably from chemical to chemical. The high-gain and low-gain curves share a similar slope for the same chemical. The intercept of the high-gain curve is lower than that of the lower-gain curve because the high-gain signal is higher than the low-gain signal for the same concentration value.

The fitting errors of the resulting concentration curves are listed in Tables 2 and 3. All concentration curves have a fitting error lower than 40%, except the high-gain curve of

vinyl chloride and the low-gain curve of ethanol. Since the high-gain curve of ethanol has a fitting error smaller than 40%, we can use the high gain curve for ethanol concentration estimation. However, vinyl chloride does not have a reliable low-gain curve. In fact, the low-gain curve error was so large that we rejected the curve while fitting it. The cause of the high fitting error of vinyl chloride curves is that the concentration change leads to a relatively small change in ion count i.e. we have a low hardware sensitivity to vinyl chloride samples. The curve slope (a) is the ratio of a change in log10(concentration) to a change in log10(signal). Therefore, a large slope means that a change in concentration leads to a relatively small change in signal. This is shown in the concentration curves of vinyl chloride in Figure 2. The relative insensitivity of the ion count to the concentration makes very difficult to distinguish different concentration using the ion count. This difficulty causes a high fitting error and consequently a high error in concentration estimation.

Testing Quantification Method

Using the established concentration curves, we tested our concentration estimation accuracy against new test measurement data. The new test data consist of over 700 elution events with over 25 chemicals, 5 different concentration values, and two or three duplicates. The test data were not used in establishing the concentration curves so they provides an independent way to gauge the accuracy of the concentration estimation using the concentration curves and the high-low gain switch decision rule. The separation between the training data and the test data also ensures that we avoid the overfitting of the concentration curve.

The test results are shown in Table 4. The listed numbers are the errors of the estimated concentration averaged over about 10 elution measurements with about 5 different concentration values for each chemical. For 25 chemicals we tested, 15 chemicals met the VCAM's concentration error budget requirement, that is, the error should be smaller than 40%. For 10 chemicals that did not meet the error budget requirement, 7 chemicals have errors between 41% and 66%. The other three chemicals (acetaldehyde, perfluoropropane, and vinyl chloride) show significant errors. For perfluoropropane and vinyl chloride, the large error is explained by the large concentration slope. The large slope means that a change in concentration is not well differentiated by the level of ion counts. The cause of the high error of acetaldehyde is likely due to coelution events with acetaldehyde during concentration curve training.

This test result also provides a way to gauge the performance of the high-low gain switch decision rule. For ethyl benzene, the dual-gain rule leads to a smaller concentration error than using only one gain data. The high gain error is 47%, the low gain error 32%, and the dual gain error 27%. This improvement is due to the built-in intelligence that the decision rule has in determining a better-quality concentration curve for a given ion-count level.

For some chemicals, the dual-gain error is the same as the lowest error between the high and low-gain. This means that one gain curve is consistently more reliable than the other gain curve. This can happen when the required concentration range of a chemical is either very low or very high. The high-gain curve would be more reliable than the low-gain curve for the very low concentration range, while the low-gain curve would be more reliable for the very high concentration range. This test result supports that our gain switch rule is successful in finding a more reliable gain curve for these chemicals.

For most of other chemicals, the dual-gain error is in between the high-gain error and the low-gain error. This indicates that the dual-gain method often successfully chooses the right gain curve but not always chooses the right one. We can improve the performance of the dual-gain rule for these chemicals by individually tuning the conditions to detect a weak signal and saturation. Currently, the conditions are global to all chemicals for simplicity.

4. CONCLUSIONS

We developed a concentration quantification method for VCAM the miniature gas chromatograph mass spectrometer instrument to be used to autonomously identify and quantify trace organic species in the International Space Station (ISS) internal air. The quantification method consists of designing a signal quantification scheme (TIC, DIC, LIC), establishing a concentration curve (relationship between the concentration and the quantified signal), and designing a dual-gain switch decision rule that chooses a more reliable gain curve to apply for a given signal level.

For this work, we measured over 1800 elution events including over 30 chemicals, 5 different concentration values per chemical, and about 5 duplicates for each chemical and concentration. About 1100 elution events were used to establish a concentration curve, and about 700 elution events were used to test the quantification method. For all chemicals except two chemicals (ethanol and vinyl chloride), the concentration curve has a fitting error smaller than 40%. We tested the concentration curve and decision rule by applying them to new independent test runs. The test campaign showed that the concentration quantification method leads to an error ranged between 14% and 66% on average for all chemicals tested except three chemicals (acetaldehyde, perfluoropropane, vinyl chloride). The three exceptions show a concentration error larger than 100%, which is partly explained by the insensitivity of the ion count to the change of concentration.

The quantification error budget allowed for VCAM is 40%. In order to reduce the error to below 40%, several approaches are considered for future work. One approach is to tune the instrument hardware gain parameters in order to increase the reliability and sensitivity of the ion count to the change of concentration. Another approach is to individually fine-tune the high-low gain switch decision rule in the quantification method for each chemical. Assurance that training and test data for concentration curve analysis does not contain coelutions would also more correctly isolate identified compounds and more accurately represent the system's capability.

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

[1] A. Chutjian, et al., "Overview of the Vehicle Cabin Atmosphere Monitor, a Miniature Gas Chromatograph/Mass Spectrometer for Trace Contamination Monitoring on the ISS and CEV," *Proceedings of the 37th International Conference on Environmental Systems*, SAE Technical Paper Series (paper number 2007-01-3150), 2007.

[2] D. L. Jan, "Environmental Monitoring Instruments: Using ISS as a Testbed for Exploration," *Proceedings of the 2007 IEEE Aerospace Conference*, Big Sky, Montana, March 2007.

[3] B. Bornstein, S. Lee, L. Mandrake, B. Bue, "Autonomous Identification and Quantification of Chemical Species with the Vehicle Cabin Atmosphere Monitor for use Onboard the International Space Station", IEEE Aerospace Conference Proceedings, March, 2008.

[4] R. G. Dromey, M. J. Stefik, T. C. Rindfleisch, and A. M. Duffield, "Extraction of Mass Spectra Free of Background and Neighboring Component Contributions from GC/MS Data," Journal of Analytical Chemistry, Vol. 48(9), 1368–1375, 1976.

[5] S. E. Stein, "An Integrated Method for Spectrum Extraction and Compound Identification from GC/MS Data," *Journal of the American Society of Mass Spectrometry*, Vol. 10, 770–781, 1999.

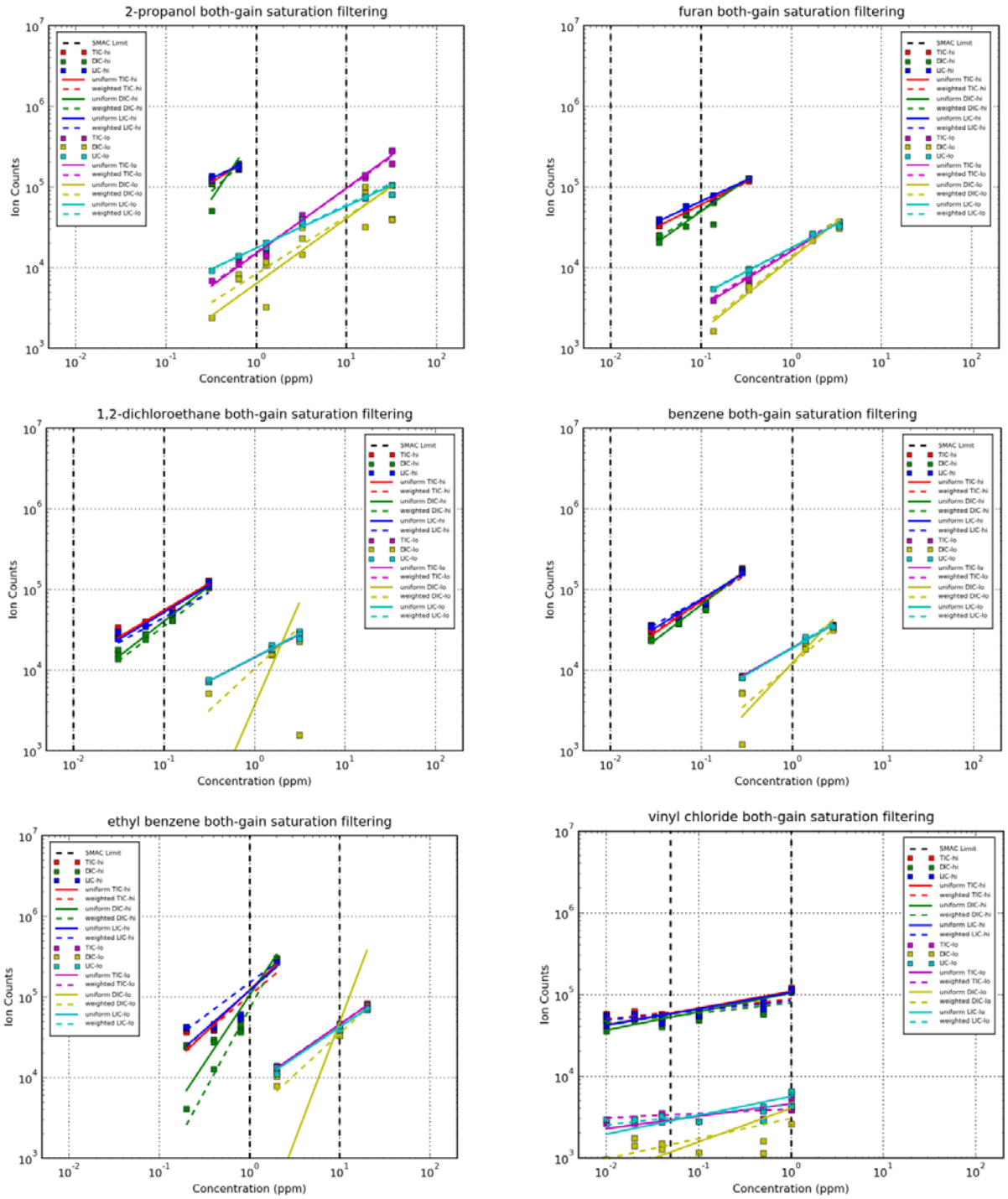


Figure 3 - Concentration curves established with three different ion-count methods (TIC, DIC, LIC) and two different fitting methods (non-weighted least square fitting and weighted chi-square fitting).

Table 1. VCAM Priority 1,2,3 species and their desired detection concentration range in ppm.

Compound	Range (ppm)	Priority
ethanol	1-10	1
acetaldehyde	0.1-3	1
acetone	0.5-5	1
dichloromethane	0.03-5	1
methanol	0.2-5	1
octamethylcyclotetrasiloxane	0.05-1	1
hexamethylcyclotrisiloxane	0.1-2	1
propylene glycol	0.5-5	1
perfluoropropane	10-100	1
1-butanol	0.5-5	2
benzene	0.01-1	2
acrolein	0.01-1	2
pentane	2-20	2
hexane	2-20	2
decamethylcyclopentasiloxane	0.1-2	2
pentanal	0.1-2	2
hexanal	0.1-2	2
ethyl benzene	1-10	2
ethyl acetate	1-10	2
2-propanol	1-10	2
ethylene (plants)	0.05-1	2
freon 113	2-10	2
furan	0.01-1	2
toluene	1-10	2
xylenes (3)	1-10	2
1,2-dichloroethane	0.01-0.1	3
alkyl amines (2)	0.5-5	3
2-butanone	0.5-5	3
4-methyl-2-pentanone	2-10	3
carbonyl sulfide	0.01-1	3
chloroform	0.02-1	3
freon 11	2-10	3
freon 12	2-10	3
isoprene	0.05-1	3
limonene	1-10	3
trimethylsilanol	0.5-5	3
vinyl chloride	0.05-1	3

Table 2. High-gain concentration curves fitted to Log-Log linear equation for several chemical species.

Chemical Name	Slope (a)	Intercept (b)	Fitting error (%)
ethanol	1.74	-8.64	33
acetaldehyde	1.09	-5.36	14
acetone	2.52	-13.77	18
dichloromethane	1.68	-9.13	20
octamethylcyclotetrasiloxane	2.57	-13.38	24
perfluoropropane	3.73	-19.36	28
benzene	1.57	-8.59	18
pentane	2.94	-16.22	25
hexane	2.13	-11.58	24
pentanal	1.44	-7.24	12
hexanal	3.01	-15.23	30
ethyl benzene	1.21	-6.05	36
ethyl acetate	1.71	-9.22	31
2-propanol	1.7	-9.1	16
freon 113	2.35	-13.58	4
furan	1.86	-9.85	9
toluene	1.44	-7.49	24
xylenes	1.14	-5.52	37
1,2-dichloroethane	2.25	-11.56	22
2-butanone	1.93	-10.25	15
4-methyl-2-pentanone	1.72	-9.36	24
chloroform	2.14	-11.59	30
freon 11	3.55	-20.26	31
vinyl chloride	5.97	-30.04	88
octane			
heptanal	2.2	-11.1	34
average over all chemicals	2.2236	-11.7388	25

Table 3. Low-gain concentration curves fitted to Log-Log linear equation for several chemical species.

Chemical Name	Slope (a)	Intercept (b)	Fitting error (%)
ethanol	3.16	-12.82	70
acetaldehyde	1.72	-6.24	22
acetone	1.46	-6.19	17
dichloromethane	1.46	-6.25	17
octamethylcyclotetrasiloxane			
perfluoropropane	3.22	-13.12	36
benzene	1.47	-6.3	13
pentane	1.68	-7.48	13
hexane	1.27	-5.42	19
pentanal			
hexanal	1.91	-7.1	9
ethyl benzene	1.34	-5.26	11
ethyl acetate	1.37	-5.9	35
2-propanol	1.26	-5.25	16
freon 113	1.07	-4.79	3
furan			
toluene	1.11	-4.32	23
xylenes	1.38	-5.17	21
1,2-dichloroethane			
2-butanone	1.6	-6.62	19
4-methyl-2-pentanone	1.2	-4.99	19
chloroform	1.39	-6.03	24
freon 11	2.04	-9.37	17
vinyl chloride			
octane	1.16	-4.51	5
heptanal			
average over all chemicals	1.6135	-6.6565	20

Table 4. Concentration estimation error by applying the concentration curves fitted with a training data set to a test data set for several chemical species. The estimation error is the difference between the estimated concentration and the real concentration relative to the real concentration. The high-gain and low-gain errors is the estimation error using the high-gain and low-gain concentration curves, respectively. The dual-gain error is given by applying the gain decision rule to determine which gain to use for concentration estimation.

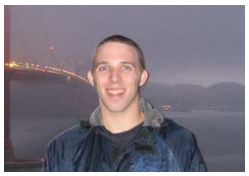
Chemical	high-gain error	low-gain error	dual-gain error
ethanol	25	185	25
acetaldehyde	494		494
acetone	55	43	45
dichloromethane	25	16	24
octamethylcyclotetrasiloxane	40		40
perfluoropropane	85	112	99
benzene	68	39	66
pentane	109	26	37
hexane	44	39	43
pentanal	23		23
hexanal	40	36	41
ethyl benzene	47	32	27
ethyl acetate	44	36	39
2-propanol	29	35	29
freon 113	59	32	35
furan	32		32
toluene	43	23	41
xylenes	45	33	42
1,2-dichloroethane	26		26
2-butanone	14	24	14
4-methyl-2-pentanone	37	36	38
chloroform	69	24	53
freon 11	24	26	28
vinyl chloride	335		335
octane		39	39

BIOGRAPHY

Seungwon Lee *is a senior member of the High Capability Computing and Modeling Group at Jet Propulsion Laboratory. She is developing flight instrument software for Vehicle Cabin Atmosphere Monitor data analysis and autonomous calibration. She is also conducting research on materials modeling and simulation, comet gas dynamics, nonlinear dynamics control, spectral retrieval, data reduction, global optimization, parallel computing, and advanced numerical algorithms. She received her Ph.D. in Physics from the Ohio State University and her M.S. and B.S. in Physics from the Seoul National University.*

Ben Bornstein *is a senior member of the Machine Learning and Instrument Autonomy group at the Jet Propulsion Laboratory in Pasadena, CA. He is the lead engineer for VCAM data analysis and its flight software implementation. Ben enjoys bringing machine learning techniques and considerable hacking (programming) skills to bear to solve problems in geology, remote sensing, bioinformatics, and systems biology. He has designed and implemented software systems for several Caltech biology labs, the Institute for Genomics and Bioinformatics (IGB) at UC Irvine, USC Children's Hospital, and JPL's Mars Exploration Rover (MER) project. He is also the inventor of and lead developer for LIBSBML, an open-source library for the Systems Biology Markup Language (SBML). Ben received a B.Sc. in Computer Science from the University of Minnesota Duluth in 1999 and is pursuing a M.Sc. in Computer Science at the University of Southern California.*

Luke Mandrake *is a member of the Machine Learning and Instrument Autonomy group at the Jet Propulsion Laboratory in Pasadena, CA. He is involved in the application of machine learning techniques to current research endeavors with various Earth-sensing satellite systems and particularly enjoys building and studying computational models of natural systems. Luke received his PhD. and M.S. in computational plasma physics from UCLA and his B.A. in physics from the University of Arizona.*



Brian Bue *is a research programmer in the Machine Learning and Instrument Autonomy group at the Jet Propulsion Laboratory, where he participates in projects involving software and algorithm development for Earth and planetary science data analysis. In the past, he has done research in computational geomorphology, automated terrain analysis, planetary image processing and scientific visualization. He received a M.S. from Purdue University in Computer Science, and Bachelor's degrees from Augsburg College in Computer Science and Mathematics. He is currently pursing a Ph.D. in Electrical and Computer Engineering at Rice University.*